Purple Group – Session 1

If you can't find it on Google...

1) If you can't find it on google, it doesn't exist (Jim)
2) Google Cell Wiki available (Brett)

Jim Sluke: If you can't find it on google, you haven't met the NIH's standard of openness
Mark Ellisman: But our datasets are so large, it needs to be separately indexed
Need annotation with googlable engines
JS: Can google parts for a 68 chevy, since manufacturers add appropriate metadata. There are more parts in a automobile than in the human genome

How do we get ranking? Crosslinks between major providers.

Jeffrey Grethe: Provide central wiki/site for tools and databases. Need community annotated list of resources

Databases should be stored with flat-file structures to enable indexing.

Shouldn't worry about data relevance until we have easy access.

Goodrelations: microformat for product data

Klaus R

Apply friend-of-friend concept to genes, with crosslinks. Allows existing social networking tools to crawl genes and work on genes.

Brett Barbaro: Include concept of 'like' button to annotate genes & networks
Lilia Iakoucheva: But biased by popularity, eg P53 connected to every gene

Need color codes for reliability of data.
Some inconsistencies come from real variability (different conditions), some from prediction failures

Blake Borgeson: Google Cell Wiki is data repository. How do we include lots of types of data, while still having structure

JS: Need a stake in the sand. For instance, many things wrong with Goodsell drawings, but still useful. Having some reference model is useful, even if known wrong.

Goodsell models would be even more useful if they were annotated with references, metadata.

JG: Googling a restaurant usually returns a yelp first hit. Thus aggregating small resources and making them available can increase access.
JS: Also improves persistence of small